Verification of Adaptive Collection for Brain Computer Interface

Mariko Matsumoto

***Abstract*— To provide speech prostheses for individuals with severe communication impairments, brain computer interfaces (BCIs) using silent speech have been studied. I proposed adaptive collection, which divided brainwaves into smaller elements and verified them, for BCIs using silent.**

This paper verified the effect of adaptive collection in comparison to the conventional method. Brainwaves were obtained when four subjects imagined vocalization. In adaptive collection, shortening time length of brainwaves for common spatial patterns was effective because the state of brainwaves changes fast when a subject imagined vocalization. As a result, using the adaptive collection with 12 ms of the time length and 20 elements for classification, the classification accuracies were improved to 87–99% and the averaged classification accuracy was improved to 93% for the pairwise classification /a/ vs. /u/ in the case of 63 channels of EEG.

***Index Terms*— Brain-computer interfaces, Brain machine interface, EEG, Common spatial patterns, Support vector machine.**

I. INTRODUCTION

We use auditory communication in daily life. Vocalization is usually used for communication. However, some disabled individuals such as amyotrophic lateral sclerosis (ALS) patients are unable to express their thoughts because a respiratory apparatus is necessary to maintain their airway. Many such individuals choose no use of the respiratory apparatuses and shorten their life time. The reason is that life without communication is called totally locked-in state (TLS) and is threatened. Brain-computer interface could be an effective technology for overcoming this problem.

For these supporting prosthetics, many studies have been conducted using methods such as P300 speller [1], steady-state visual evoked potentials (SSVEP) speller [2], SSVEP cursor controller [3], and near infrared spectroscopy (NIRS) [4]. However, in SSVEP, the capability for eyes to move and to be united in a focus is needed. In the method using hemodynamic response, e.g., NIRS, users must train in the mode of imagining calculations or fast songs for detection. The methods described above necessitate training of skills that users have never developed in daily life. The classification of silent speech is a simple method that requires no special training.

Dasalla (2009) classified imagined vocalization of vowels using scalp electrodes, common spatial pattern (CSP) filtering, and support vector machine with Gaussian kernel [5]. The classification accuracies (CAs) were 56–72% for the pairwise classification /a/ vs. /u/ in the case of 64 channel EEG measurement.

In the earlier paper, I proposed adaptive collection [6]. Using the adaptive collection, the CAs were improved to 73-92% in the similar condition.

However, the effect has not been sufficiently verified. In this paper, I divided the effect of the adaptive collection into three parts, (1) time reduction of input signals for CSPs, (2) selection of valid elements for classification, and (3) a large number of collection in order to verify the effect of the adaptive collection.

II. PROCEDURE FOR PAPER SUBMISSION

A. Subjects

Four 21–24-year-old male subjects were native speakers of Japanese who were right-handed, as assessed by the Edinburgh Inventory [9]. No participant had any neurological disorder or noteworthy health problem. Experiments were conducted in accordance with the Declaration of Helsinki. Informed consent was obtained from all subjects.

B. Experiments

Each subject was pasted electrodes and seated comfortably in an armchair with eyes closed to avoid the influence of visual activation. The subjects were coached beforehand and had rehearsed with actual movements a few times to ensure correct task execution. The subjects were then asked to imagine voice production while remaining silent and immobilized. Two tasks were conducted: the fixed order task and the random order task. The tasks used sound commands generated by a personal stereo device (Walkman NW-E053; Sony Corp.). Subjects, while hearing them through earphones, were instructed to perform the following tasks.

Manuscript received November 18, 2014

Mariko Matsumoto, Information Science, Hiroshima City University, Hiroshima, Japan,

1) Fixed order task

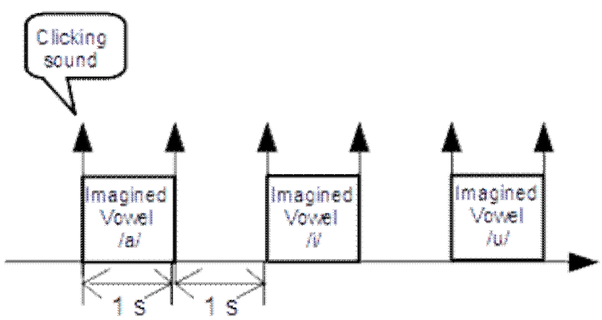
Subjects were instructed to imagine the voice production of one of vowels /a/, /i/, /u/, /e/, and /o/ in fixed order for one second following one second for rest. The onset and ending of the imagined vocalization were signaled to the subjects using clicking sounds (Fig. 1(a)).

One trial stream consisted of 5 vowels × 13 times for about 2.2 min. Subjects performed the experimental set four times. In all, 52 epochs were obtained for each vowel. Consequently, 260 trials were obtained for each subject. We designate this batch of data as 260 epochs.

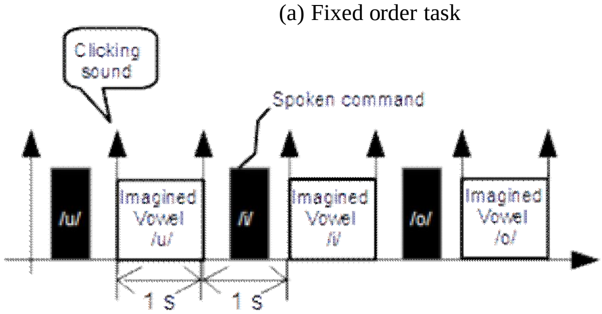
2) Random order task

Subjects were instructed to imagine the voice production (imagined vocalization) of a vowel that was the same as the last spoken command for one second following one second for rest. The spoken command expressed one of the vowels, /a/, /i/, /u/, /e/, or /o/ in randomized order. The onset and ending of the imagined vocalization were signaled using clicking sounds (Fig. 1(b)).

To avoid the influence of auditory evoked potentials, the interval between spoken commands and onset of the imagined speech was set to 200 ms or more. In all, 50, 45, 52, and 52 epochs were obtained, respectively, for each vowel for subjects 1, 2, 3, and 4. To clarify, in task /a/, for instance, subjects imagined speech production of /a/ for one second, while they remained silent and immobilized. The ways for other vowels are as the same as above.



(a) Fixed order task



(b) Random order task

Fig 1: Experimental protocols

C. Recording

EEG signals were recorded using an electroencephalograph (Neurofax EEG-1100; Nihon Kohden Corp.) and 128 channel Modular EEG-Recording Caps (Easycap GmbH) with a sampling rate of 1000 Hz. The recorded EEG data were zero-phase band-pass filtered at frequencies of 0.1–300 Hz to avoid anti-aliasing noise and to remove any low-frequency baseline shift.

I recorded brain waves using 111 electrodes. For the verification, I reduced the number of electrodes from 111 to 63. The remaining electrodes for calculation were therefore 63 electrodes. For reference, two electrodes were attached on the right and left ears. One electrode was set below an eye to detect unwanted eye movement and artifacts. However, no artifact rejection algorithm was used for this study.

D. Data processing

Data processing was performed using software (MATLAB; The MathWorks Inc., Natick, MA). Using a decimation filter with cutoff frequency of 62.5 Hz or 125 Hz, the recorded EEG data were decimated from 1000 Hz sampling to 125Hz or 250 Hz sampling after filtering respectively. 62.5 Hz or 125Hz was respectively determined from 125 Hz or 250Hz of sampling frequency by the Nyquist theorem. Epochs were extracted in reference to the stimulus onset. The duration was one second. 50 epochs were extracted for Subject 1, for a total of 250 epochs. The total numbers for Subjects 2, 3, and 4 were, respectively, 225, 260, and 260 epochs.

III. Method

The method of Dasalla (2009) used 500ms interval brainwaves of 64 channels. For classification, it selected four signals which were constituted of upper two and lower two vectors of CSP. 56-72% of the classification accuracies were obtained [5].

In my earlier reports, I proposed adaptive collection. It applied CSP filters for 100ms portions of 64 brainwaves. It selected the most effective 20 elements and combined their results. The classification accuracies were improved to 73-92% [6]. Table 1 shows the summary.

Table 1: Summary of method

Authors	Method
Dasalla et al. [5]	It used a CSP filter using 64 brainwaves of 500ms. The 500ms was from onset to 500ms. Classification used combining four results, which were associated with upper two and lower two vectors of CSP. The classification method was SVM. The cut off frequency of low pass filter was 62.5 Hz.
Matsumoto [6]	It used adaptive collection (AC). Where ten CSP filters used ten *t*-elements, each of which consists of 100ms duration of 63-ch brainwaves. The AC divided output signals of the CSPs into 630 elements and used 20 elements for classification. The 20 elements were selected by validation. The classification method was SVM. The cut off frequency of low pass filter was 125 Hz.

In order to verify the effect of the adaptive collection, I evaluated difference from the method of Dasalla (2009).

There are three different points between these methods.

(1) Split in the time axis of brainwaves

(2) Selection of effective elements

(3) Large number of elements to be used for the identification

In order to verify the effect of above three points, this paper compared the following cases;

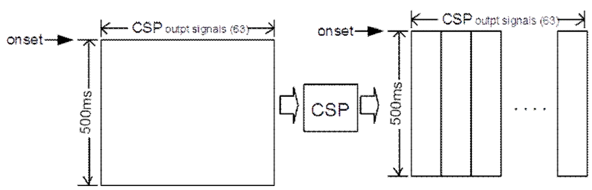
a) No selection : It used a CSP filter using 63 brainwaves of 500 ms. The 500 ms was from onset to 500 ms (Fig. 2 (a)). The classifier used combining *M* SVM results, which were associated with upper *mu* and lower *ml* vectors of CSP. *M* is the sum of *mu* and *ml*. *mu* is equal to *ml*. In this case, cut off frequency of low pass filter was 62.5 Hz. The sampling speed was 125 Hz.

In the random order task, the point at M = 4 means "Pseudo Dasalla (2009)". The number of electrodes of "Pseudo Dasalla (2009)" is only different from the condition of "Dasalla (2009)".

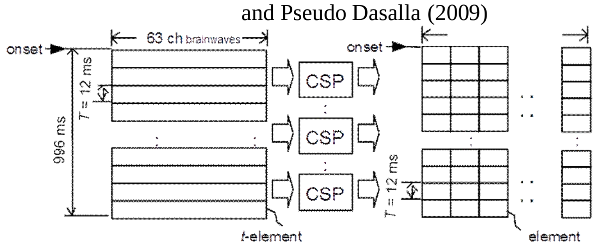
b) Simple selection : It used a CSP filter using 63 brainwaves of 500 ms. The 500 ms was from onset to 500 ms (Fig. 2 (a)). It evaluated output signals of CSP and selected the most effective *M* signals for classification using SVM. In this case, cut off frequency of low pass filter was 62.5 Hz. The sampling speed was 125 Hz.

c) The adaptive collection (AC) with various time length and the number of collection : This paper verified various values of the time length *T* and the number of collection *M*. *T* was set to 12 ms, 48 ms, 100 ms, 200 ms, and 500 ms. *M* was set to 4, 10, 20, 30, and 40. When *T* was 12 ms, the total number of elements was 5,229 and it used signal from onset to 996ms (Fig. 2 (b)). In this case, cut off frequency of low pass filter was 125 Hz. The sampling speed was 250 Hz.

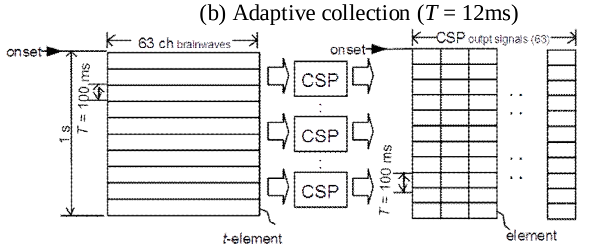
The AC with T = 100 ms and M = 20 is similar to the matsmoto (2013) (Fig. 2 (c)).



(a) No selection, simple selection, and Pseudo Dasalla (2009)



(b) Adaptive collection (*T* = 12ms)



(c) Adaptive collection (*T* = 10ms) and matsumoto (2013)

Fig 2: Element generation.

A. *Adaptive Collection*

Adaptive collection (AC) enables the use of suitable time window and vectors of CSP for classification.

In other words, it adaptively uses effective output signals of CSPs for classification to improve classification accuracies (Figs. 3).

The AC consists of the *t*-element generation, the element generation, the evaluation, and the combination and decision (Fig. 3).

For batch data calculation, two epochs are used as test data and the remaining epochs are divided evenly into training data and validation data every an iteration. Then various combinations of test data, training data, and validation data are used iteratively. Eventually all epochs are used as test data. The validation data are used for evaluation whereas test data are used for classification. For example, when the number of epochs is 52 for each vowel, number of test data is two, number of training data and evaluation data is 51 for pairwise classification.

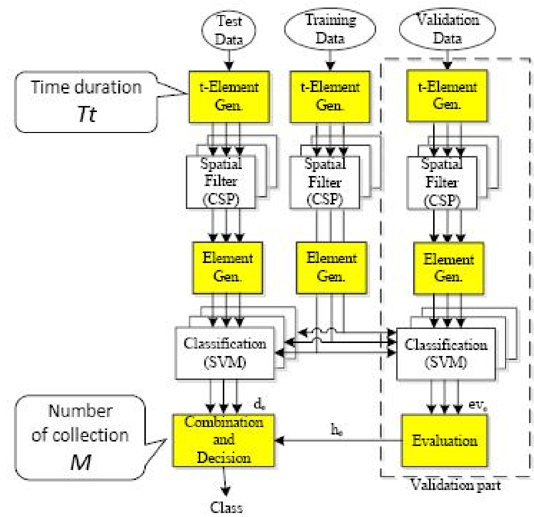


Fig 3: Flow generation.

1) *t*-Elements Generation

Data is divided in the time domain into *t*-elements. Each *t*-element consists of brainwaves with time duration of *T*, which is set to 12 ms, 48 ms, 100 ms, 200 ms, and 500 ms. The *t*-element label is *t-e* = (*t*), where *t* denotes time label (t = 1, 2, ··· , 10, when T = 100 ms) (Fig. 2 (c)). The *t*-elements are made for application of CSPs.

2) Element Generation

After CSPs, each data is divided into elements. Each element consists of one output signal of a CSP with time duration of *T*. The element label is *e* =(*s*, *t*), where *s* denotes signal label (*s* =1, 2, ··· , 63) (Fig. 2 (b), (c)).

3) Evaluation

To ascertain the suitable elements for classification, I evaluate the performance related to the elements per subject and vowel combinations. The evaluation uses the validation data and the training data, and these are strictly isolated from the test data. The reliability coefficient was set to one

when its classification accuracy was included in the top M, and is otherwise zero. M is various.

4) Combination and Decision

To use suitable elements from the evaluation results, the combination and detection part outputs the class using the following formula.

$$\text{class} = f\left(\frac{\sum_{e=1}^{N_e} h_e r_e}{\sum_{e=1}^{N_e} h_e}\right) \quad (1)$$

Therein, r_e is the classification result related to element e, e.g., class 1 or 2, Ne is the total number of elements, and $f(\cdot)$ is the decision function. The parenthesis above represents the average of classification results related to the top M elements.

As a result, AC selects suitable time duration and spatial feature of brainwaves for classification because the output signals of CSP are related to eigenvectors.

B. Common Spatial Patterns (CSP)

Spatial filter enables single trial detection without averaging of multiple trials [5]. Detailed descriptions of the CSP principle can be found in reports of studies by Müller-Gerking et al. [7] and Ramoser et al. [8]. To describe them briefly, given two groups of EEG time series data (e.g., tasks to classify /a/ and /u/), I designate each epoch as a matrix $E_g^{t,i}$ in which the rows and columns of E respectively denote electrodes and samples. t is the time label, i is the epoch label, and g is the group label. I then compute normalized covariance matrices for the epochs of each group and each element and average them such that

$$C_g^t = \frac{1}{m}\sum_{i=1}^{m} \frac{E_g^{t,i}\left(E_g^{t,i}\right)^T}{trace\left(E_g^{t,i}\left(E_g^{t,i}\right)^T\right)} \quad (2)$$

where m is the number of trials in group g. The two resultant matrices are summed to produce a composite covariance matrix C_c^t, which is then factored into its eigenvectors such that the following apply.

$$C_c^t = C_1^t + C_2^t \quad (3)$$

$$C_c^t = V_c^t \lambda_c^t V_c^{t\,T} \quad (4)$$

Therein, V_c^t is a matrix of eigenvectors λ_c^t is a diagonal matrix of eigenvalues. I then calculate a linear transformation called a “whitening transformation”.

$$\mathbf{W}^t = \sqrt{\lambda_c^{t\,-1}} V_c^{t\,T} \quad (5)$$

It equalizes the variances in eigenspace. The whitening transformation is then applied to the original two covariance matrices.

$$S_g^t = W^t C_g^t W^{t\,T} \quad (6)$$

$$S_1^t = U^t \lambda_1^t U^{t\,T} \quad (7)$$

Thereby, the transformation renders their eigenvectors U^t equivalent and their eigenvalues and summing to 1, with the diagonal elements of 1 ordered in ascending order. Finally, I define a projection matrix $P^t = \left(U^{t\,T} W^t\right)^T$, where the columns of P^{t-1} are the common spatial patterns. They can be regarded as time-invariant EEG source distribution vectors during an element; then I decompose each EEG epoch such that

$$Z_g^{t,p} = P^t E_g^{t,p} \quad (p \not\subset i) \quad (8)$$

The resultant feature vectors of $Z_g^{t,p}$ are optimized for discrimination of the two groups, where p, the epoch label for the test data, is isolated from i of Eq. (2). The exception is only itself for test data. The exception is test data and itself for validation data. In that way, I calculated CSPs for each subject, vowel combination, and time label and used them.

In Fig. 3, the input of the spatial filter (CSP) is $E_g^{t,p}$ in Eq. (8) and $E_g^{t,i}$ in Eq. (2) and the output is $Z_g^{t,p}$ in Eq. (8).

C. Support Vector Machine with Gaussian kernel (SVM-G)

The support vector machine classifier is a classifier algorithm that seeks an optimal hyperplane as a decision function in a high-dimensional space by use of training data [10, 11].

Adopting a radial basis function (RBF) with Gauss kernel enable non-linear classification [11]. The kernel function of support vectors is

$$\mathrm{K}(x, x_j) = e^{-\frac{\|x-x_j\|^2}{2\sigma}} \quad (9)$$

where σ is a parameter related to variation of the training data. In this paper, parameter was determined through a cross-validation of the validation data. Using “SVM and Kernel Methods Matlab Tool box” of an SVM software package, I applied SVMs with Gaussian kernels for pairwise classification.

IV. Results

Fig. 4 and Fig. 5 show the averaged classification accuracies (CAs) over all subjects for parewise classifications of /a/ vs. /u/ in the case of random order tasks and fixed order tasks, respectively. The legend symbols "No selection" and “Simple selection ” respectively denote no selection and Simple selection in the methods section. A point of "Simple selection" at M = 4 means Pseudo Dasalla (2009).

“ AC (T=12ms)”, “AC (T=28ms)”, “AC (T=48ms)”, “AC (T=100ms)”, and “AC (T=200ms)” respectively denote results obtained by the adaptive collection when the time duration T was 12 ms, 28 ms, 48 ms, 100 ms, and 200 ms. The number of collected elements, M was set to 4, 10, 20, and 40. The number of channels was 63.

In Fig. 4, except of a point at M = 4 in “AC (T=48ms)”, the shorter the time duration is, the better the classification accuracies are. This is because when M is small, results has large variation.

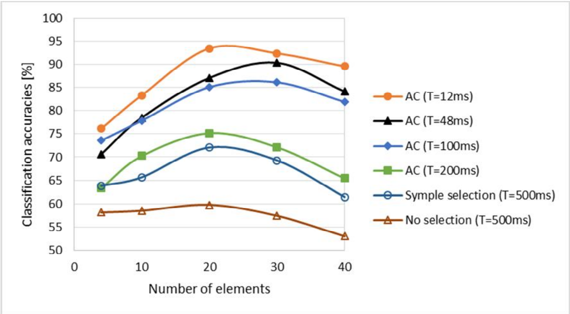


Fig 4: Classification results for various time duration (random order tasks).

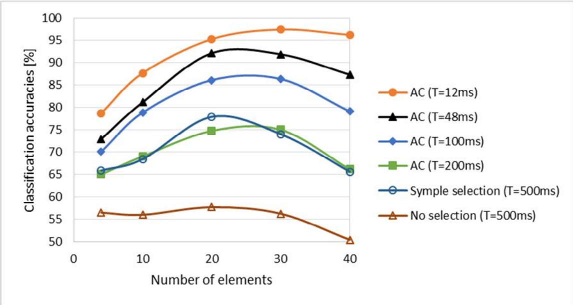


Fig 5: Classification results for various time duration (fixed order tasks).

In Fig. 5, except of a point at M = 20 in “AC (T=200ms)”, the shorter the time duration is, the better the classification accuracies are. This is because cut off frequencies of low pass filters for “Simple selection” and “AC (T=200ms)” are different.

V. Discussion

In order to verify the effect of the adaptive collection, I simulated using various time durations and various number of collections. Some major conclusions derived from the results are the following.

First, in comparing "No selection" and "Simple selection", you can find the effect of selection. The results shows that classification accuracies (CAs) of the simple selection are better than that of the no selection. The simple selection did not always select the vector associated with the maximum eigenvalue, while the no selection selected them. From these results, it can be said that the selection is effective because the vector associated with the maximum eigenvalue is not always valid for selection.

Second, result shows the effect of time duration. Classification accuracies of short time duration were better than those of long time duration except of a few results. For examples, at M = 20, a result in “AC (T=12ms)” was 93% while that in “AC (T=200ms)” was 75%. This is because the state of brainwaves change rapidly when a subject imagines vocalization. The effect is that the short time duration make it possible to follow the fast change of brainwaves.

Third, results show the effect of number of combination. When you focus on legend symbol "No selection" in Fig. 4, a result of M = 4 is worse than that of M = 20. It means that the greater the number of collected elements M is, the more stable the CAs are. However, the point at M = 40 was the weaker than the point at M = 20. The CAs of M = 40 was close to the chance level because it included the poor elements.

If you pay attention to "Simple selection" or “AC” in Fig. 4, the trend was clearer than that of "No selection".

When the time duration was shorter, the CAs at M = 40 were better. For the short time duration, there are many effective elements, because the total number of elements became large, for example, whereas 5,229 for T = 12 ms, 315 for T = 200 ms.

Fourth, if you doubt that the results were occurred from not imagined vocalization but auditory stimuli in random order tasks, the results of fixed order tasks show the evidence of the influence of imagined voice, because the fixed order task used only clicking sound. The results of fixed order tasks (Fig. 5) show the same trend as those of random order tasks (Fig. 4).

Fifth, the above results proofed the reason of the improvement of matsumoto (2013) from Dasalla (2009). In Fig. 4, the point of legend symbol "No selection" at M = 4 is “Pseudo Dasalla (2009)” and the point of legend symbol "AC (T=100ms)" at M = 20 is matsumoto (2013).

There were three effective points, (1) split in the time axis of brainwaves, (2) selection of effective elements, and (3) large number of elements to be used for the identification.

In addition, using adaptive collection of T = 12 ms and M = 20, 93% of the CA was obtained in the pairwise classification /a/ vs. /u/ for 63 channels. It is better than that of the earlier report. The shorter time length achieved the better performance.

In this study, I evaluated pairwise classifications and did batch processing. Multiple classifications and online processing remain as subjects for future study.

VI. Conclusion

For silent speech decoder, the adaptive collection has three effects, (1) shortening brainwaves for calculating CSP enables to follow the fast change of the silent speech, (2) selection of effective elements for the identification improves the CAs, and (3) large number of elements to be used for the identification enables performance to be stable. In addition, using the adaptive collection with T = 12 ms and M = 20, 93% of averaged CA was obtained in the classification /a/ vs. /u/ for 63 channels. It is better than the earlier results. The shorter time length achieved the better performance for decoder of silent speech.

References

[1] E. Donchin, K. Spencer, and R. Wijesinghe, “The mental prosthesis: assessing the speed of a P300-based brain–computer interface,” IEEE Trans. Rehabil. Eng., vol. 8, no. 2, pp. 174–179, 2000.

[2] M. Cheng, X. Gao, S. Gao, and D. Xu, "Design and implementation of a brain–computer interface with high transfer rates," IEEE Trans. Bio-Medical Eng., vol. 49, no. 10, pp. 1181–1186, 2002.
[3] L.J. Trejo, R. Rosipal, and B. Matthews, "Brain–computer interfaces for 1-D and 2-D cursor control: designs using volitional control of the EEG spectrum or steady-state visual evoked potentials," IEEE Trans. Neural Systems Rehabil. Eng., vol. 14, no. 2, pp. 225–229, 2006.
[4] M. Naito, Y. Michika, K. Izawa, Y. Ito, M. Kiguchi, and T. Kanazawa, "A communication means for completely locked-in ALS patients based on changes in cerebral blood volume measured using near-infrared light," IEICE Trans, Inf. Syst., E90-D(No.7), pp.1028–1037, 2007.
[5] W C.S. DaSalla, H. Kambara, M. Sato, and Y. Koike, "Single-trial classification of vowel speech imagery using common spatial patterns," Neural Networks, vol. 22, no. 9, pp. 1334–1339, 2009.
[6] M Matsumoto, "Silent speech decoder using adaptive collection," 19th IUI, pp. 73–76, 2014,
[7] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," Clinical Neurophysiology, vol. 100, pp. 787–798, 1999.
[8] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," IEEE Transactions on Rehabilitation Engineering, vol. 8, no. 4, pp. 441– 446, 2000.
[9] A. Rakotomamonjy, "Variable selection using SVM based criteria," The Journal of Machine Learning Research Archive, vol. 3, 3/1/2003, pp. 1357–1370, 2003.
[10] B.E. Boser, I.M. Guyon, and V.N. Vapnik, "A training algorithm for optimal margin classifiers," In D. Haussler, editor, Fifth Annual ACM Workshop on COLT, pages 144–152, Pittsburgh, PA, ACM Press, 1992.
[11] F. Asano, M. Kimura, T. Sekiguchi, and Y. Kamitani, "Classification of movement-related single-trial MEG data using adaptive spatial filter," IEEE ICASSP, pp. 357–360, 2009.